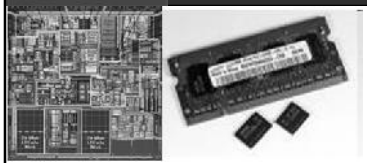
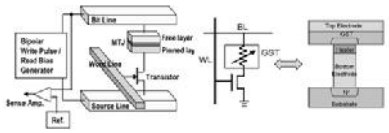
UCSB





MPSOC 2015

Memory That Never Forgets:

Exploiting nonvolatility for architecture design

Yuan Xie
UCSB

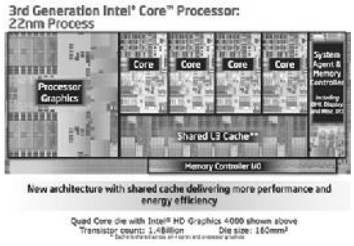
Scalable and Energy-efficient Architecture Lab (SEAL)

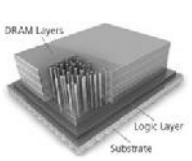
http://seal.ece.ucsb.edu/ SEAL@UCSB

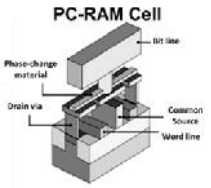
Emerging Computing Architectures –

HPC (mobile, enterprise, embedded) computer design is more fluid now than in the past two decades.

- **Heterogeneous processing**
  - Latency tolerant cores
  - Throughput cores
  - Special purpose hardware (e.g., AES, MPEG, RND, Machine Learning)
- Memory
  - 2.5D and 3D Stacking
  - HMC, HBM, WIDEIO2, LPDDR4, etc
  - New devices (PCRAM, ReRAM)
- **Interconnects**
  - Various workloads
  - Scalable topologies
  - Optical, wireless interconects
- **Storage**
  - Active storage
  - Non-traditional storage architectures (key-value stores)

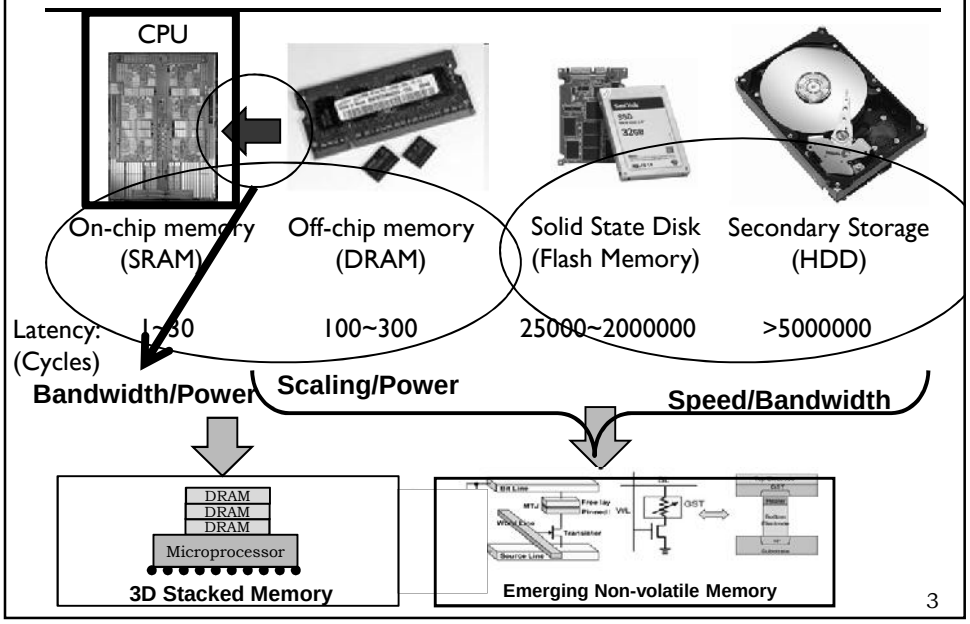






2

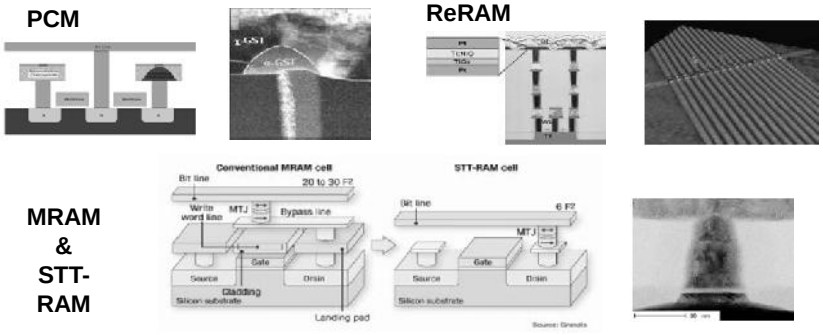
Challenges in today's computer system: **Memory Wall**



3

Emerging Non-Volatile Memories

- Emerging NVM are promising candidates as replacement for SRAM/DRAM/NAND flash



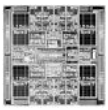
Comparison of Memory Technologies

Jeffrey Vetter, ORNL
Robert Schreiber, HP Labs
Trevor Mudge, University of Michigan
Yuan Xie, Penn State University

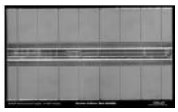
	SRAM	DRAM	NAND Flash	PCRAM	STTRAM	ReRAM
Data Retention	N	N	Y	Y	Y	Y
Cell Size (F^2)	50-200	4-6	2-5	4-10	8-40	4
Minimum F demonstrated (nm)	14	25	16	20	28	27
Read Time (ns)	< 1	30	Slow	Medium	Fast	Medium
Write Time (ns)	< 1	50	Slow	Medium	Fast	Medium
Number of Rewrites	10^{16}	10^{16}	10^4-10^5	10^8-10^{10}	10^{15}	10^8-10^{12}
Read Power	Low	Low	High	Low	Medium	Medium
Write Power	Low	Low	High	High	Medium	Medium
Power (other than R/W)	Leakage	Refresh	None	None	None	Sneak
Maturity						

http://ft.ornl.gov/trac/blackcomb

Opportunities



On-chip memory (SRAM, MRAM, ReRAM)



Off-chip memory (DRAM, PRAM, ReRAM)

Solid State Disk (Flash Memory, PRAM, ReRAM)

Secondary Storage (HDD)

- Leveraging NVM for last level on-chip caches
  - Performance and power improvement
- Combining NVM and NAND flash for hybrid SSD
- Leveraging NVM to improve the reliability of the on-chip memory
- Leveraging NVM for instant power-on/power-off

6

What are the most unique features for NVM?

	SRAM	eDRAM	PCM	STT-RAM	ReRAM
Cell Area (F^2)	**50-200**	**30-40**	**4-10**	**6-20**	**<=4**
Read Power	**Low**	**Low**	**Low**	**Low**	**Low**
Write Power	**Low**	**Low**	**High**	**High**	**High**
Other Power	**Leakage Current**	**Refresh Power**	**None***	**None***	**None***
Write Endurance	**Unlimited**	**Unlimited**	**10^8**	**10^{12} ~10^{15}**	**10^{11}**

Non-volatile !

7

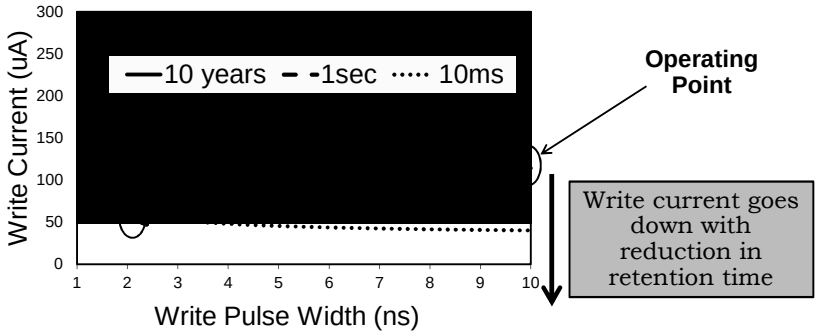
Reduce Retention Time (Nonvolatility)

For cache/memory, years of *data-retention time* may not be required.

- *Trade-off* retention time for lower write latency
- Opportunity
  - Architecting "Volatile STT-RAM" Caches
- Advantage
  - Performance and Energy Benefits

8

An example



Retention Time of STT-RAM	Write Latency @ 2 GHz
10 Years	22 cycles
1 second	12 cycles
10 millisecond	6 cycles

Ref: "Cache Revive: Archi-tecting Volatile STT-RAM Caches for Enhanced Performance " DAC 2012

How to leverage Non-volatile Features?

A fast check-pointing device

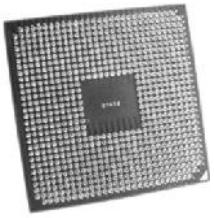
- Xiangyu Dong, Naveen Muralimanohar, Norm Jouppi, Richard Kaufmann, Yuan Xie "**Leveraging 3D PCRAM Technologies to Reduce Checkpoint Overhead for Future Exascale Systems**." SC 2009
- Ping Chi, Cong Xu, Tao Zhang, Xiangyu Dong, Yuan Xie "**Using Multi-Level Cell STT-RAM for Fast and Energy-Ecient Local Checkpointing**" ICCAD 2014, IEEE/ACM William J. McCalla ICCAD Best Paper Award.

Fault Resiliency for Exascale System

- Microprocessor becomes unreliable
  - Process scaling, voltage scaling, soft error, NBTI, ……
- Even assuming socket MTTF remains constant
  - system MTTF = socket MTTF / number of socket

1 socket

Socket MTTF = 5 years



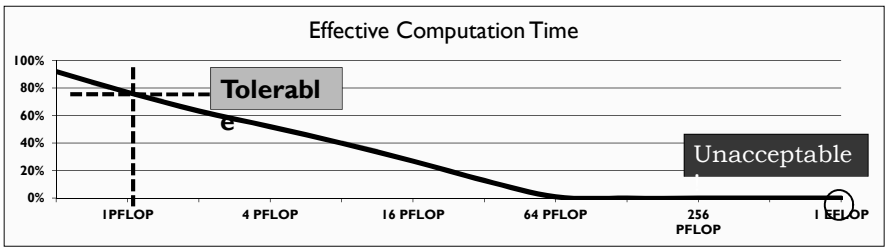
Exascale ~100,000 socket

System MTTF = 26 minutes

11

Checkpoint / Restart

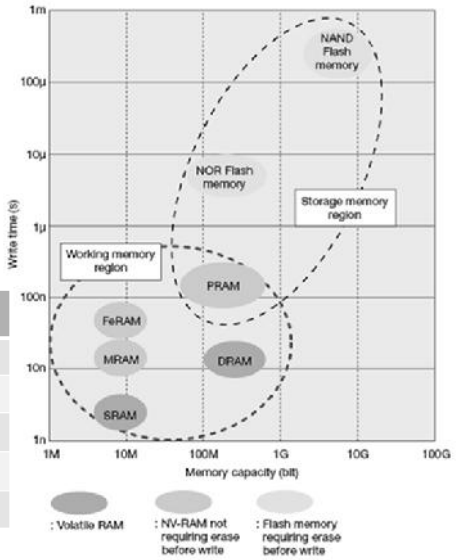
- Checkpoint / Restart is the state-of-the-art
  - HDD or SSD as the checkpoint storage
  - HDD peak bandwidth: ~100MB/s, SSD: ~800MB/s
  - BlueGene/L: 12 mins to take a checkpoint
    - Equivalent to 8% performance loss
- Scale to exascale …



PCRAM – A Good Candidate

- PCRAM is 2 orders faster than flash
- PCRAM has 3 orders higher endurance than flash
- Good candidate for local checkpoint

	HDD	NAND Flash	PCRAM
Cell size	-	4-6F^2	4-6F^2
Read time	~4ms	5us-50us	10ns-100ns
Write time	~4ms	2ms-3ms	100-1000ns
Stanby power	~1W	~0W	~0W
Endurance	10^{15}	10^5	10^8

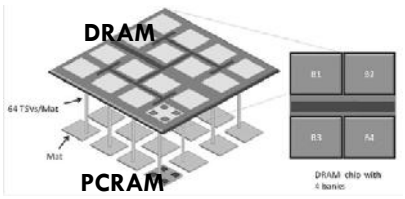


Courtesy: Motoyuki Ooishi

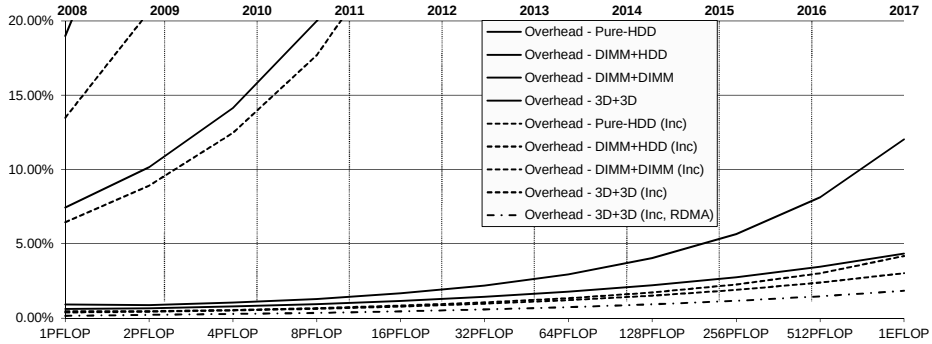
How to Integrate PCRAM

- Option 1: PCRAM DIMM
  - Deploy PCRAM on DIMM
  - Possible local bandwidth ~10GB/s (SSD bandwidth ~800MB/s)

- Option 2: 3D PCRAM
  - Deploy PCRAM directly on top of DRAM
  - Possible local bandwidth ~2.5TB/s (DIMM bandwidth ~10GB/s)



Our Projection

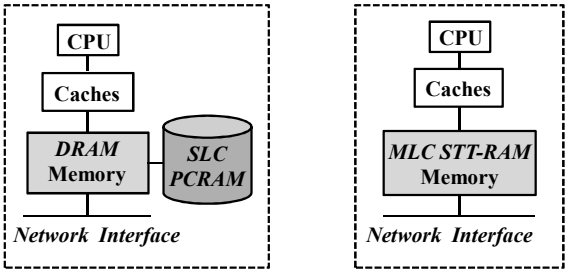


- If 10% checkpoint overhead is considered as acceptable
  - Pure-HDD is not going to work soon.
  - DIMM+HDD: till year 2011
  - DIMM+DIMM: till half exaFLOPS
  - 3D+3D: beyond exascale!

Ref: "Leveraging 3D PCRAM Technologies to Reduce Checkpoint Overhead for Future Exascale Systems." SC 09

Using MLC STT-RAM

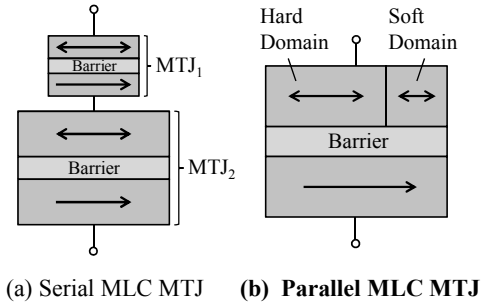
- leverage MLC STT-RAM as check-pointing device
  - MLC provides inherent multi-versioning opportunity
  - relatively simple write operations vs. MLC PCRAM



- Ping Chi, Cong Xu, Tao Zhang, Xiangyu Dong, Yuan Xie "**Using Multi-Level Cell STT-RAM for Fast and Energy-Ecient Local Checkpointing**" ICCAD 2014, IEEE/ACM William J. McCalla ICCAD Best Paper Award.

MLC STT-RAM

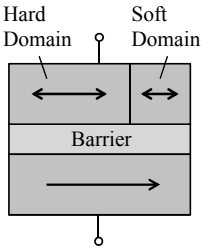
- two types
  - serial MLC
    - two MTJs
    - + easier to fabricate
  - parallel MLC √
    - two domains
    - + higher TMR ratio
    - + smaller switching current density



(a) Serial MLC MTJ **(b) Parallel MLC MTJ**

Parallel MLC

- soft bit: stored in the soft domain
  - easier to flip
  - require a small switching current
- hard bit: stored in the hard domain
  - harder to flip
  - require a larger switching current

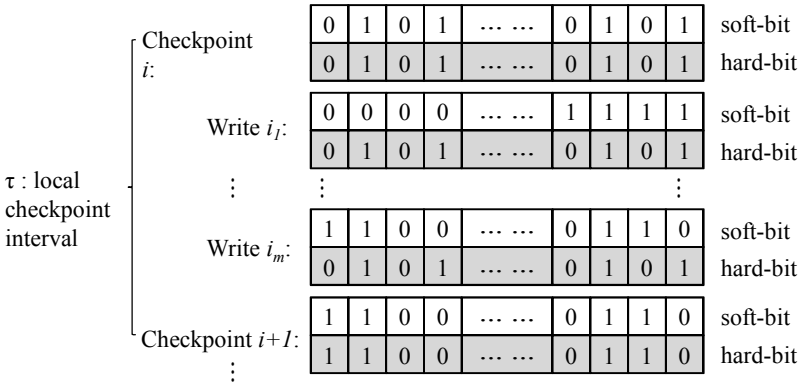


Parallel MLC MTJ

- four resistance levels: R00, R01, R10, R11
  - MSB -> hard bit
  - LSB -> soft bit

The proposed mechanism

- intra-cell local checkpointing
  - soft bit: working data; hard bit: checkpoint data



How to leverage Non-volatile Features?

Persistency Support

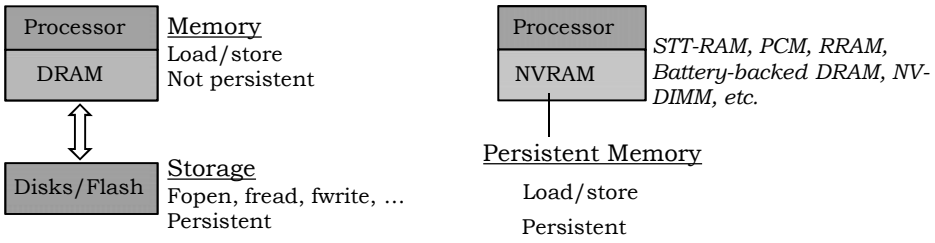
- Rethink cache/memory hierarchy design

  J. Zhao et al., "Kiln: Closing the performance gap between systems with and without persistence support," MICRO'13. Best Paper Honorable Mentioned

- Explore new memory scheduling schemes

  J. Zhao et al., "FIRM: Fair and High-Performance Memory Control for Persistent Memory Systems," MICRO'14.

Persistent Memory

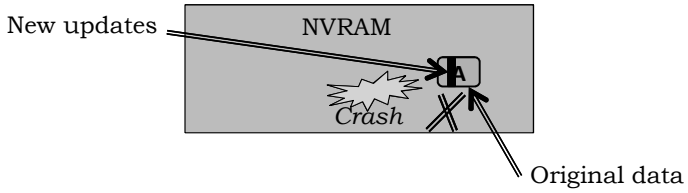


- Allow access to permanent data through a memory interface
- Demonstrated much faster than using traditional memory/storage stack

Ref: [Condit+ SOSP'09, Volos+ ASPLOS'11, Coburn+ ASPLOS'11, Venkataraman+ FAST'11]

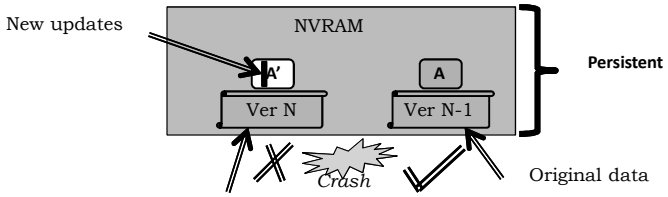
21

Persistence Support: Multiversioning



22

Persistence Support: Multiversioning



- Log (logging)
- Temporary buffer (copy-on-write)

Source of overhead –
Two writes with each one data update

Memory traffic increased ~2X!

[Volos+ ASPLOS'11, Coburn+ ASPLOS'11, Condit+ SOSP'09, Venkataraman+ FAST'11]

23

What Can be Leveraged From Hardware

Software's view of memory system

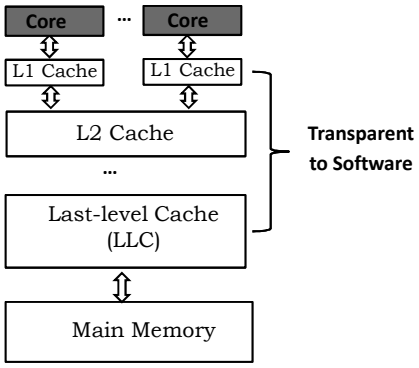
- **A flat address space**
- **Pages**

Page belonging to process

Page not belonging to process

Hardware's view of memory system

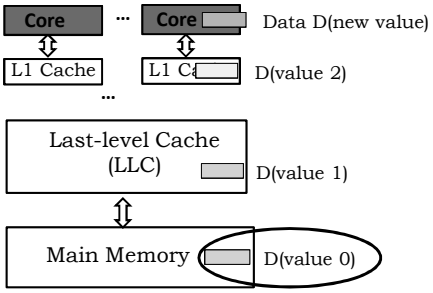
- **Not flat, a hierarchy**
- **Cache lines**



24

Multiversioning: Leveraging Caching

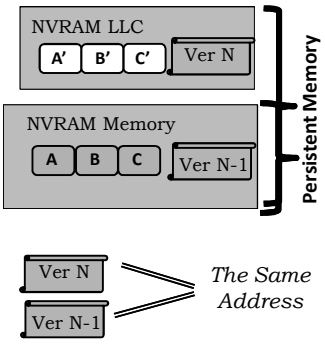
- How does a cache hierarchy work?
- A multiversioned system by nature!



25

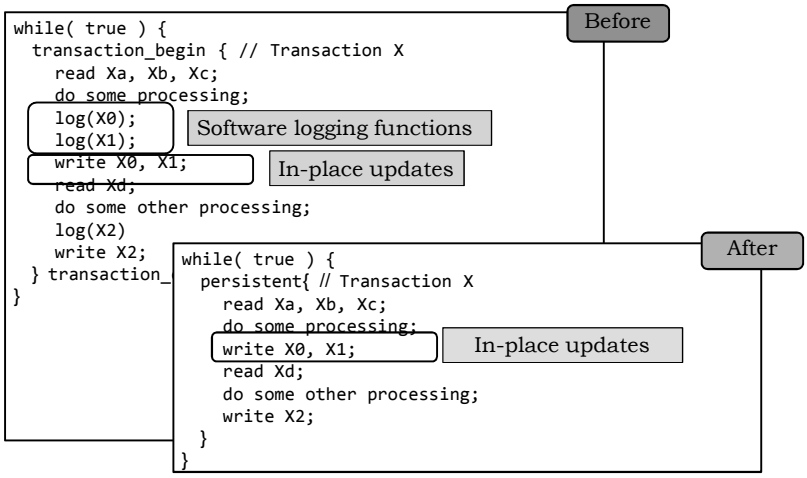
Multiversioning: Leveraging Caching

- How does a write-back cache work?
- A multiversioned system by nature
- Design a persistent memory hierarchy
  - Updates directly overwrite original data
  - No need for logging or copy-on-write



26

Code Examples
Before and After Using Our Design



27

How to leverage Non-volatile Features?

Non-Volatile Processor for Energy-Harvesting

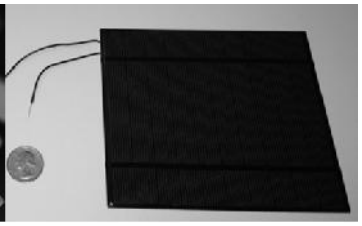
- Kaisheng Ma, Yang Zheng, Shuangchen Li, Karthik Swaminathan, Xueqing Li, Yougpan Liu, Jack Sampson, Yuan Xie, Vijay Narayanan "**Architecture Exploration for Ambient Energy Harvesting Nonvolatile Processors.**" HPCA 2015. Best Paper Award.

- Kaisheng Ma, Shuangchen Li, Xueqing Li, Yougpan Liu, Jack Sampson, Yuan Xie, Vijay Narayanan "**Nonvolatile Processor Architecture Exploration for Energy Harvesting Applications**" , IEEE Micro, 9/2015

Ambient Energy Harvesting Sources

RF Energy

Thermal Energy

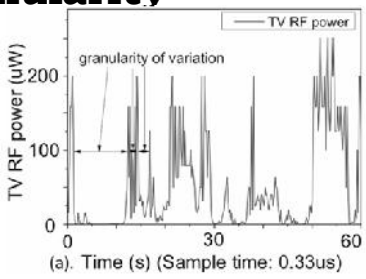


Solar Energy

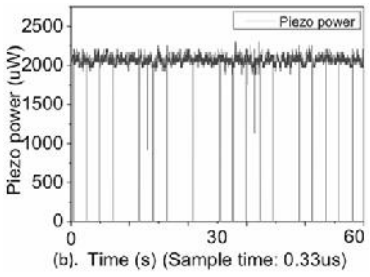
Piezoelectric/Vibration Energy

29

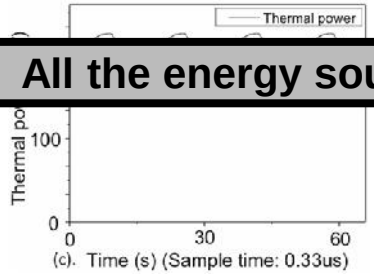
Signal Magnitude, Variability and Granularity



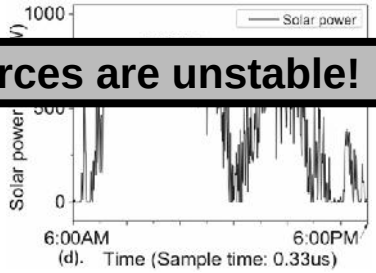
TV RF: Unstable & large variation



Piezo: Periodical power on/off



Thermal: Relatively stable but low

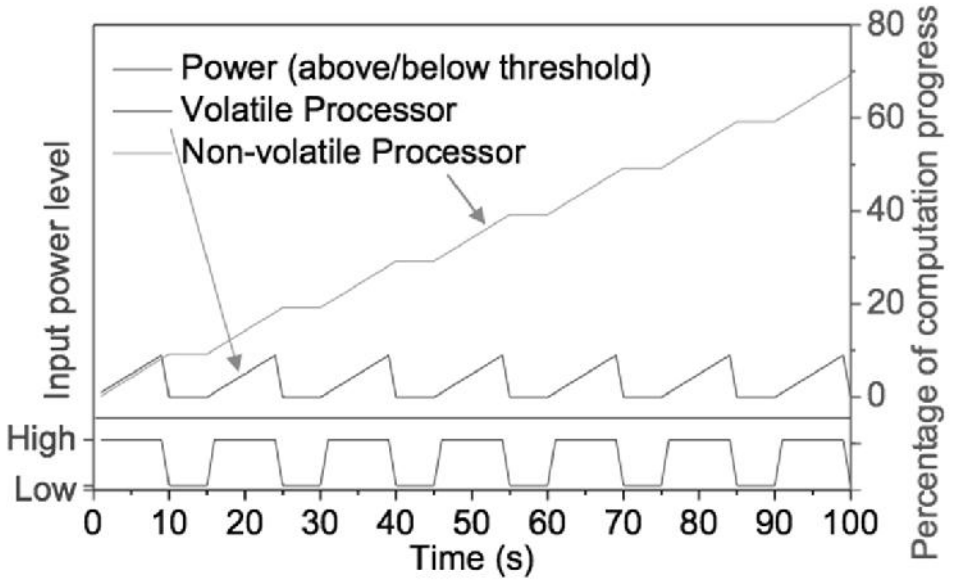


Solar: Time and environment dependent

All the energy sources are unstable!

30

Volatile or Nonvolatile ?

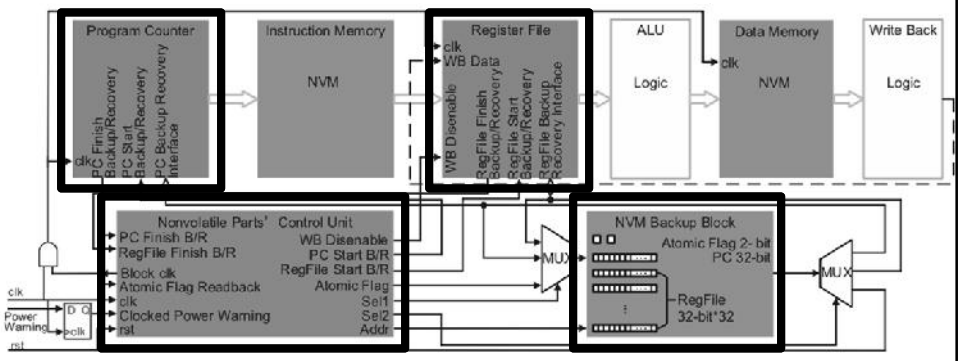


31

New Optimizing Targets

- **Backup (recovery) energy:**
  - Energy consumption for backing up(restoring) processor states, as the overhead of the system.
- **Forward progress:**
  - The completed percentage of task running on a processor, depending on not only CPI, but also power profiles.

From Volatile to Nonvolatile

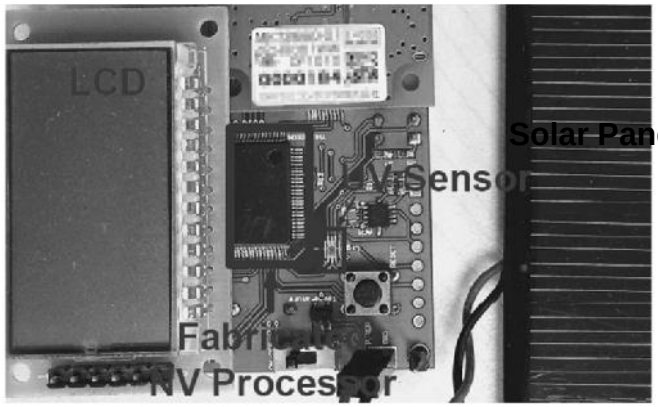


➢ Kaisheng Ma, Yang Zheng, Shuangchen Li, Karthik Swaminathan, Xueqing Li, Yougpan Liu, Jack Sampson, Yuan Xie, Vijay Narayanan "**Architecture Exploration for Ambient Energy Harvesting Nonvolatile Processors.**" HPCA 2015. Best Paper Award.

33

Prototyping System

- Intel 8051 Core, using 0.13 µm ROHM CMOS-ferroelectric hybrid process. The PC and all RegFiles are FeRAM-based Flip-Flops
- The NV processor-based system is interfaced to a solar power panel and a UV sensor



Conclusion

- Emerging NVM has many benefits for new memory architecture design
  - High-density, lower standy-power, no refresh power
- Write overhead and endurance are challenges to be addressed
  - Write delay/energy overhead: hybrid cache, wear-limiting
  - Endurance: Wear-leveling
- Nonvolatility provides new opportunities for novel designs.
  - New checkpointing device
  - Persistent memory support
  - NVM processor
  - Any other new opportunities?

Thank you!

Q&A